

# ECHO: Foundational Middleware for a Science Cyberinfrastructure

Robin Pfister (robin.g.pfister@nasa.gov)  
Keith Wichmann (wichmann@gst.com),  
Beth Weinstein (beth.weinstein@nasa.gov)

MODIS Science Team Meeting  
March 22, 2005



## ECHO is...



---

Is an Enabling Framework that allows interoperability among diverse and distributed, data, service and client systems

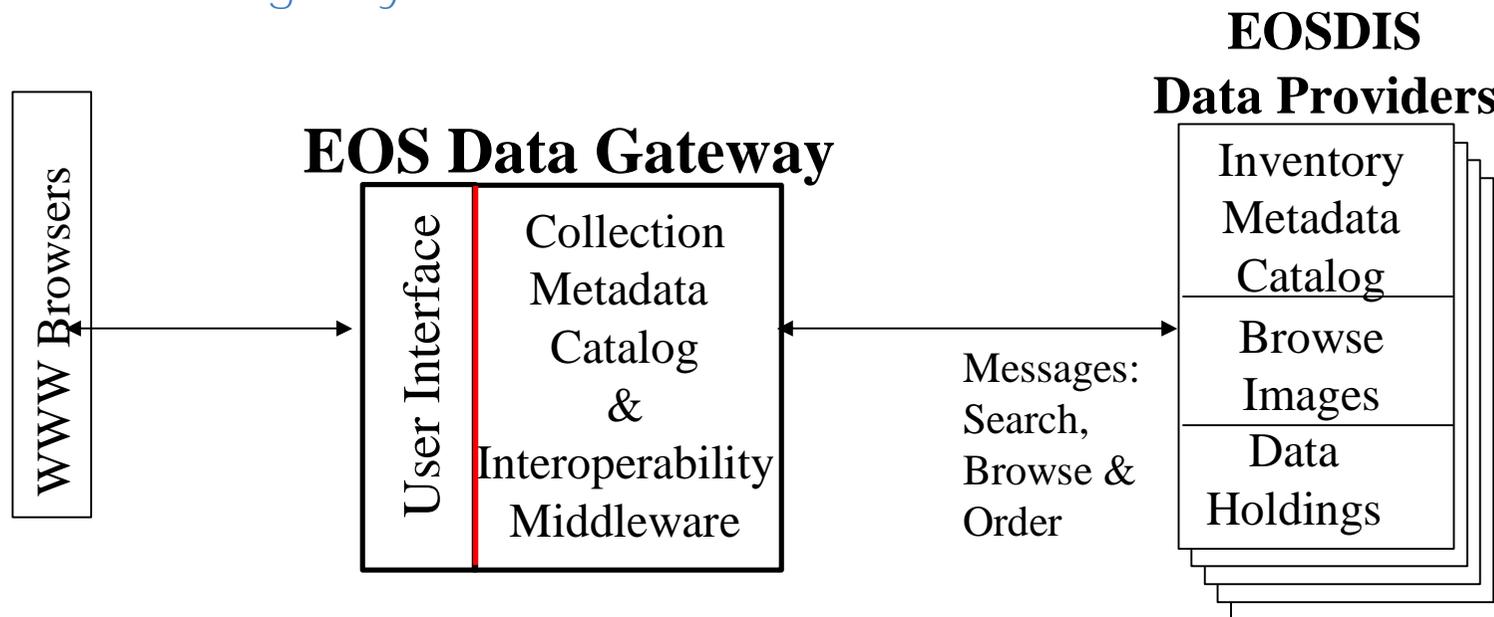
Is a metadata clearinghouse and order broker. In the near-future will also be a granule-level service broker

Is an Open System: Enabling a Collaborative Community

Users of Earth Science Data were confined to a single user interface built as a one-size-fits-all solution

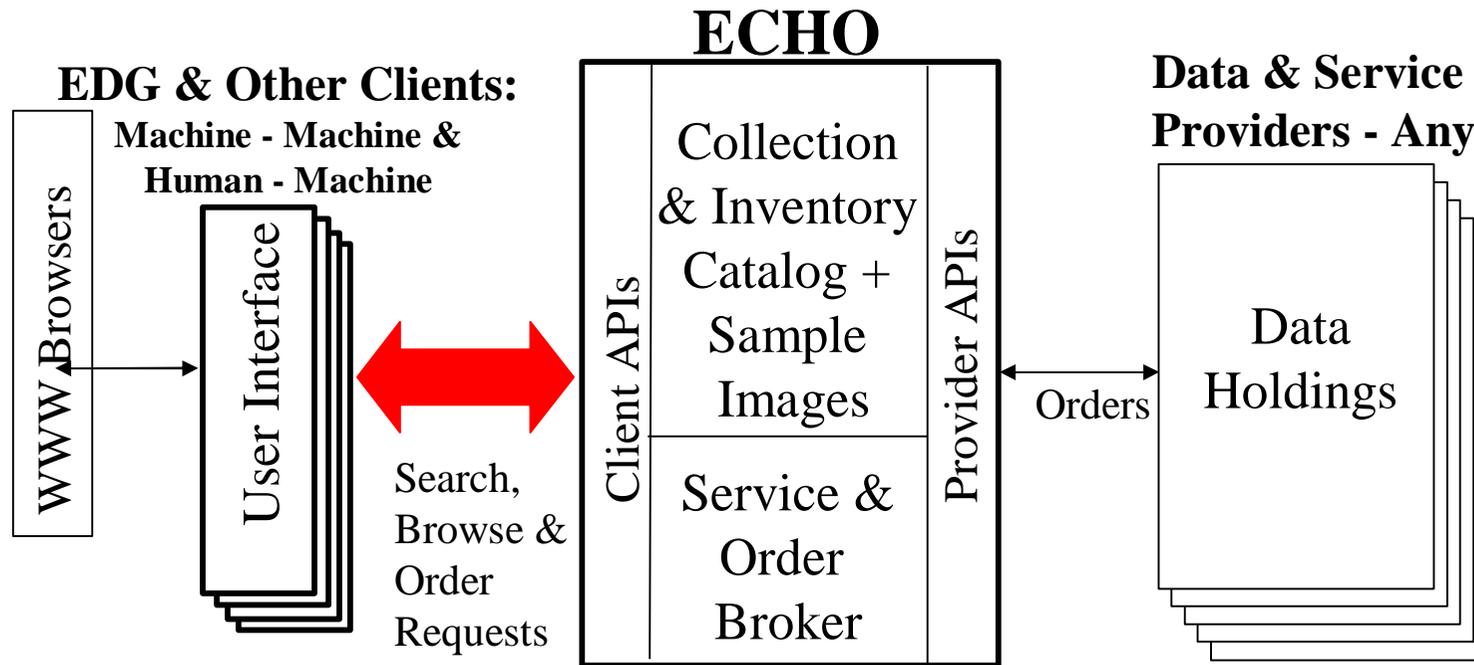
Search and retrieval data access paradigm, searches were executed at the archives. Many factors contributed to slow or non-performance  
 Scientists spent a significant amount of time performing data access activities

NASA was paying a single contractor to develop all needed functionality into a single system



# Background

Individuals and communities can share data, services and tools.  
 Scientists save time by using tools tailored for their data access needs, cutting out unnecessary steps in the data access process  
 Tools can be developed by anyone, *saving NASA funds and increasing quality* through more competition



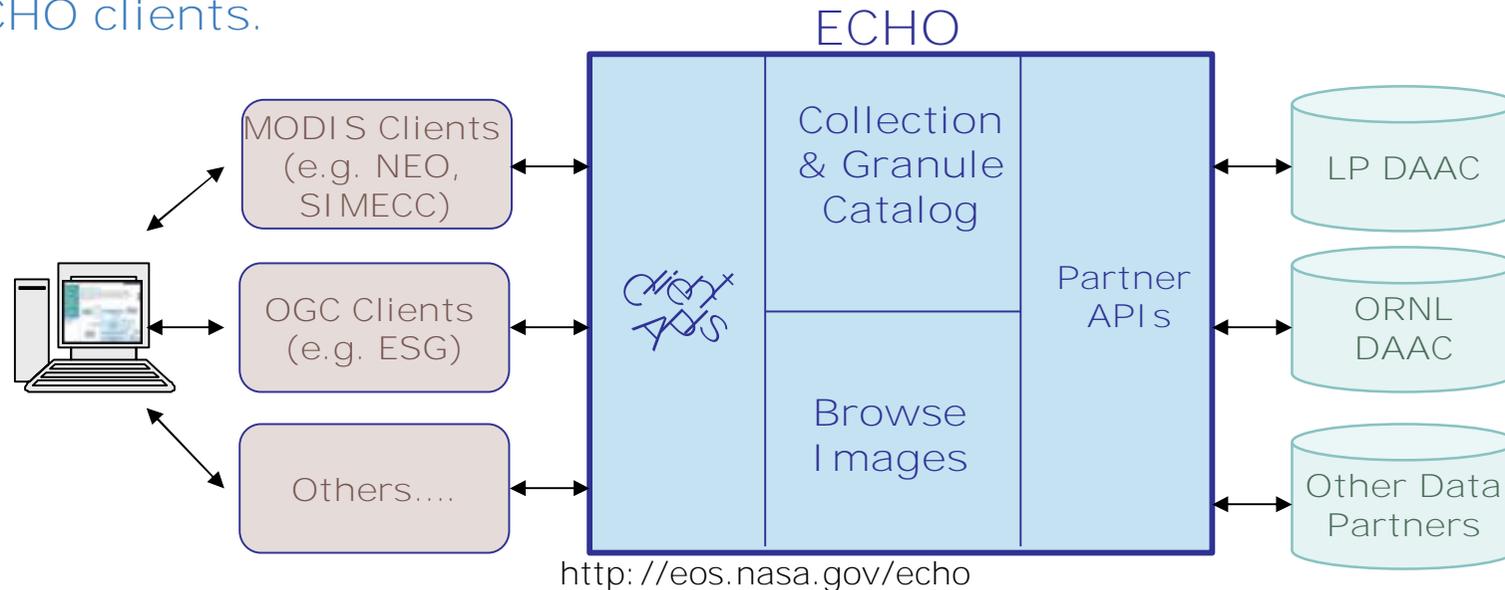
# ECHO Context

ECHO acts as middleware between resource (data or service) partners and client partners.

- o Clearinghouse of spatial, temporal, numeric and textual Earth Science metadata
- o Order Broker
- o User and provider account service
- o Services clearinghouse and broker (future)
- o Capabilities are in line with (and expected to be driven by) GEOSS.

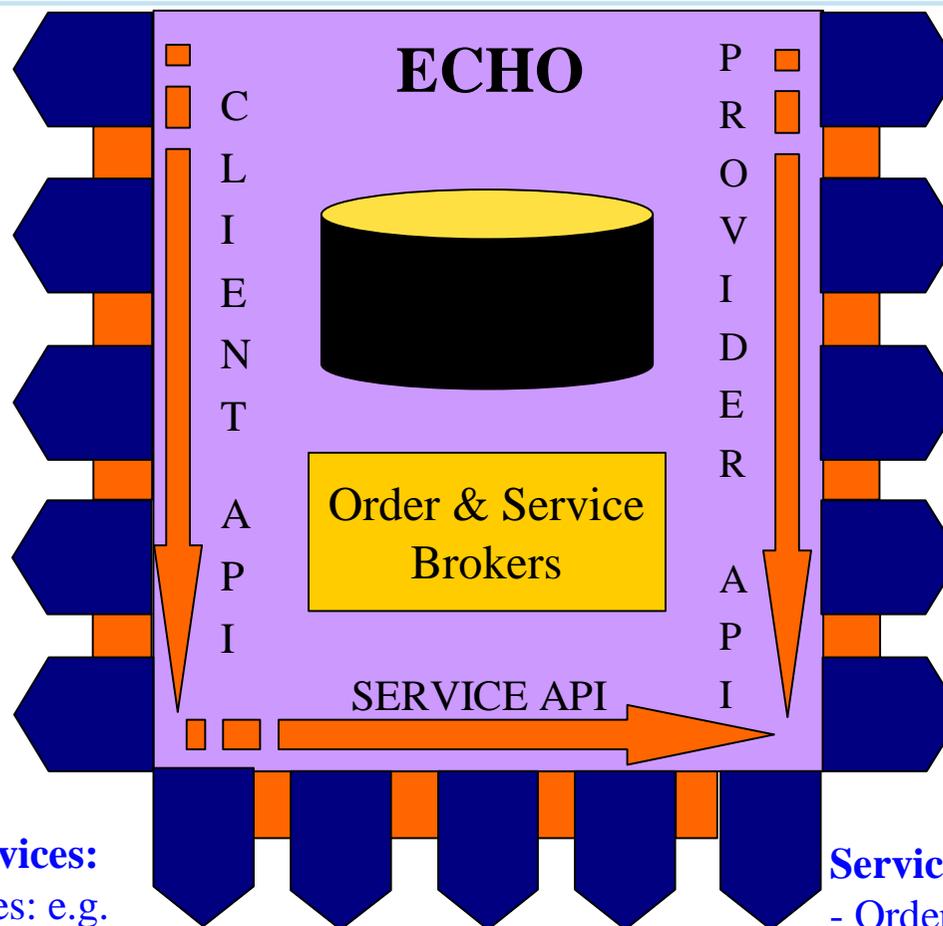
Data Partners provide information about their data holdings, and client partners develop software (either machine-machine or human-machine) to access this information.

End users who want to search ECHO's metadata must use one of the ECHO clients.



### Client Extensibility

- Graphical User Interfaces (applets, active pages, etc.)
- Machine-to-Machine Interfaces



### Data Extensibility

- New Data Partners
  - Can establish policies for their own data
- New Collections/Data Types
- Access Mechanisms

### Types of services:

- Data Services: e.g. collection, subsetting
- Search Services: e.g. gazetteer, thesaurus (future)
- Administrative Services: e.g. accounting (future)

### Service Extensibility

#### Service Views:

- Service View
- Data View

<http://eos.nasa.gov/echo>

### Service interactions:

- Order Options
- Advertised
- Brokered (future- ECHO 8.0)
- Orchestration/Chaining (future)



## Why use ECHO?

---

The multi-organizational content of ECHO provides a broadened set of resources to a growing number of Earth science applications and interdisciplinary research efforts.

- ∅ The more resources provided, the richer the set of available resources.

ECHO streamlines access to digital data and materials and brokers orders and other services from Clients to Data Partners.

- ∅ More data access paradigms and approaches are supported.
  - Human-Machine and Machine to Machine
  - “Navigation and discovery” and “Search and retrieval”
  - Subscriptions
- ∅ Virtual seamless access from the end user’s perspective.
- ∅ Because scientists will have access to clients tailored to their needs, valuable time can be saved and refocused on research.



## Why use ECHO? (cont.)

---

ECHO puts the control of the resources in the hands of the providers of those resources.

- ECHO has tools for providers to manage and control access to their contributed resources.
- Establish the best spatial search approach (details in back-up slides) for their data (orbital search, geodetic, or Cartesian)

Improved availability over previous architecture (details in back-up)

ECHO provides tracking services for both the Data and Client Partner.

In the future, new technologies can be infused using the services APIs.

Back-up slides also illustrate:

- Recent performance improvement. Performance improvement is an ongoing effort (more details on availability and performance improvements in back-up slides)



## Current State of ECHO

---



Current Operational Version: 5.5.3

Version 6.0 is being externally tested

Version 7.0 is ready for external test

Version 8.0 development is getting started

Information on future releases ECHO Web site.



# Client API Improvements

---



Client Workshop (March 10-11) kicked off 2005 as the "year of the client".

- o Prioritize API areas for development of the Client Libraries and the Reference client.
- o Reference Client will provide a view with 'best practices' code that can be copied and pasted into other applications.
- o The workshop also encouraged component sharing among the existing client providers.

Additionally, We've repackaged the complete set of toolkit source code, java documentation, and libraries.

- o This repackaging fills in all the remaining "holes" in the toolkit library
- o Included is a sample demonstration client- this demo client was always there - it is updated to reflect the 6.x changes.



# ECHO Holdings Report



## Operational Holdings Summary as of 3/1/2005

	Collections	Granules	Browse
Public	1233	35,178,394	5,308,581
Restricted	176	6,483,460	2,047,530
Total	1,409	41,661,854	7,356,111



# ECHO Data Partners

---



## Operational

- ∅ ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center)
- ∅ SEDAC (Socioeconomic Data and Applications Center)
- ∅ LP DAAC (Land Processes EOS Core System (ECS) DAAC)
- ∅ GES DAAC (GSFC Earth Sciences ECS DAAC)
- ∅ ASF DAAC (Alaska SAR Facility DAAC)

## In Test

- ∅ LaRC DAAC (Langley Research Center ECS DAAC)
- ∅ NSIDC DAAC (National Snow and Ice Data Center ECS DAAC)

## In Development

- ∅ PO.DAAC (Physical Oceanography DAAC)
- ∅ SSC SDP (NASA Stennis Space Center Science Data Purchase)
- ∅ GES DAAC V0 (GSFC Earth Sciences Version 0 DAAC)



# ECHO Client Partners

---

## Operational

- Mercury EOS – ORNL client
- Power User Interface – script for bulk ordering
- SIMECC (Simple MODIS ECHO Client) – MODIS data search and order

## In Test

- WIST – EOS Data Gateway using ECHO
- ASF ECHO Client - Alaska SAR Facility client
- WISRD (Web Interface for Searching, Subsetting, Stitching, Resampling, Regridding, and Reformating Data) – NSIDC Client to search for swath, scene, and gridded data
- MODIS Website – Client via existing MODIS website and ORNL's shopping cart
- Data Validation User Interface – Client for MODIS Land Data Validation Team

## Under development

- Annoterra Version2 – Link between EOS newsfeeds, GCMD, and ECHO granules
- SNOWI-E – NSIDC client
- NEO - NASA Earth Observations client
- Invasive Species Data Service
- OGC/NSDI Client Adaptor - will allow any OGC/NSDI compliant client to plug in. Earth-Sun Gateway (ESG) is an example )

There are plans for other clients (MODIS rapid response, Adaptive Sensor Fleet)



## Client Partner Support

---

The ECHO Operations Team (ECHO Ops) is the point of contact for direct interaction between ECHO, its Partners, and end users

ECHO Ops is responsible for the operation and maintenance of the ECHO operational and partner test systems

ECHO Ops general support for Partners includes:

- Assisting Partners in understanding the ECHO API and DTDs
- Ingest management and accounting - weekly summaries and metrics on all ingest jobs
- Problem tracking and resolution
- Advertising Data Partner holdings and availability of new datasets
- Promoting and engaging new Client Partners

Reference client in progress (best practices code reuse)

Improve website content to provide "front office" for client developer support

- Improve overview materials
- Create sense of thriving project and community
- Reorganize developer reference materials into developer center

Develop multilevel documentation to ease use of APIs



# Contact Information

---



## ECHO Ops

- ◊ [echo@killians.gsfc.nasa.gov](mailto:echo@killians.gsfc.nasa.gov)
- ◊ 301-867-2071

## ECHO website

- ◊ <http://eos.nasa.gov/echo>
- ◊ Holdings Summary, Upcoming Functionality, APIs and DTDs
- ◊ Real-time status

## ECHO mailing lists

- ◊ echo-all, echo-status, echo-client, echo-data

## ECHO Technical Committee (ETC) Meetings

- ◊ Tuesday @ 3:00pm ET
- ◊ Contact ECHO Ops for dial-in information



## DVUI Demonstration

---



Data Validation User Interface Proof-of-concept  
Illustrates ECHO can support very different search and access paradigms from EDG  
Resulting from this effort were improvements to ECHO that are being released in upcoming iterations  
This is an example of something that couldn't be done in the old heritage architecture  
Potential re-users of this concept (or code) need to make sure that the data they need for their application are made visible by the providers of those datasets to the potential users of the client (this was a major reason DVUI wasn't able to go operational before)



# Backup

---



**DVUI: Coincidence Evaluation**

Go To: Reference Granule Search

Validation Site:

Day  Night  Both  No DN Info

Date Range: 2002/05/13 - 2002/05/24

Distance from Validation Site (km): 0 - 200

Cloud Coverage Percentage: 0 - 10

Include Granules Without Cloud Info

Reference Granule: E15C:L70RWR8.002:2006518042 **20 Results Found** [View Results](#)

No.	Granule ID	Dataset ID	Size (MB)	Start Date	Center Lat.	Center Lon.	Distance	% Cloud Cover	Day/Night	Browse	Shopping Cart
13	E15C:L70RW...	LANDSAT-7 L...	0.0	2002-05-24	38.9025	-78.4993	40	1	DAY	<a href="#">Browse</a>	<input type="checkbox"/>
14	E15C:L70RW...	LANDSAT-7 L...	0.0	2002-05-17	37.4747	-77.3872	86	1	DAY	<a href="#">Browse</a>	<input type="checkbox"/>
15	E15C:L70RW...	LANDSAT-7 L...	0.0	2002-05-16	38.9088	-79.3964	123	0	NIGHT	<a href="#">Browse</a>	<input type="checkbox"/>
16	SC:MIL2ASLS...	MISR Level 2 ...	138.816	2002-05-15	39.377747	-79.63188	0		DAY	<a href="#">Browse</a>	<input type="checkbox"/>
17	SC:MIL2ASLS...	MISR Level 2 ...	97.169	2002-05-24	39.377743	-78.08681	0		DAY	<a href="#">Browse</a>	<input type="checkbox"/>
18	SC:MIL2ASLS...	MISR Level 2 ...	53.5628	2002-05-17	39.377743	-76.54171	0		DAY	<a href="#">Browse</a>	<input type="checkbox"/>



## Architecture Overview: System Drivers

---



Design for Change - with ever-changing technology, science needs and policies we needed to build a highly-evolvable system

Present an open API for organizations to connect their own data or service or to build user interfaces for specialized views of these resources

Make it easy for providers of Earth Science data and services to participate in the system

Provide searches that respond quickly

Broker orders for both data and services

Minimize operational costs

Build upon advances in industry and use e-commerce systems as a model

Build a system that can be scaled up to handle large numbers of requests



## Access Control Lists

---

ECHO supports allowing the Data Partner to restrict visibility of their Collection or Granule metadata, as well as the ability to order

Since ECHO contains 10s of millions of records, granules to be access controlled are determined based on a set of rules

- Fixed time rules (i.e. hide data acquired between September 11th, 2001 and September 30th, 2001)
- Relative time rules (i.e. prevent orders for data for 30 days after it is produced)
- Restriction flag setting
- Specific collection or granule

Access controls are evaluated at the last possible moment so that changes are taken into account

Providers block access to everyone, and then grant access to groups

ECHO provides a separate function for managing what users are in a group

PUMP (Provider User Management Program) is a GUI tool for managing access controls and other Data Partner related functions



## About Spatial Data

---



ECHO leverages Oracle's spatial capabilities to perform spatial search and to represent Partner metadata  
Both collections and granules can be spatially represented

- ∅ If a collection or granule represents the entire Earth, there is no need to spatially index it
- ∅ ECHO is asking Providers to set a spatial keyword to "Global" in order to indicate this condition
  - Version 5.5 will have a special Global flag for the Partner to set

Oracle 9i version 2



# Cartesian, Geodetic and Orbit Models

---



## Cartesian Model

- ∅ The Earth is represented by a rectangle (-180, -90, 180, 90)
- ∅ Spatial objects cannot cross the International Date Line
- ∅ Spatial objects cannot cross the poles
- ∅ Polygons can be as large as the whole Earth

## Geodetic Model

- ∅ The Earth is represented by a spheroid
- ∅ Spatial objects may cross the International Date Line
- ∅ Spatial objects may cross the poles
- ∅ No single polygon may have an area equal to or larger than half of the Earth's area

## Orbit Model







# Services Provided For Searching

---



## Spatial Search

- o Cartesian
- o Geodetic
- o Orbit-based

## Temporal Range Searches

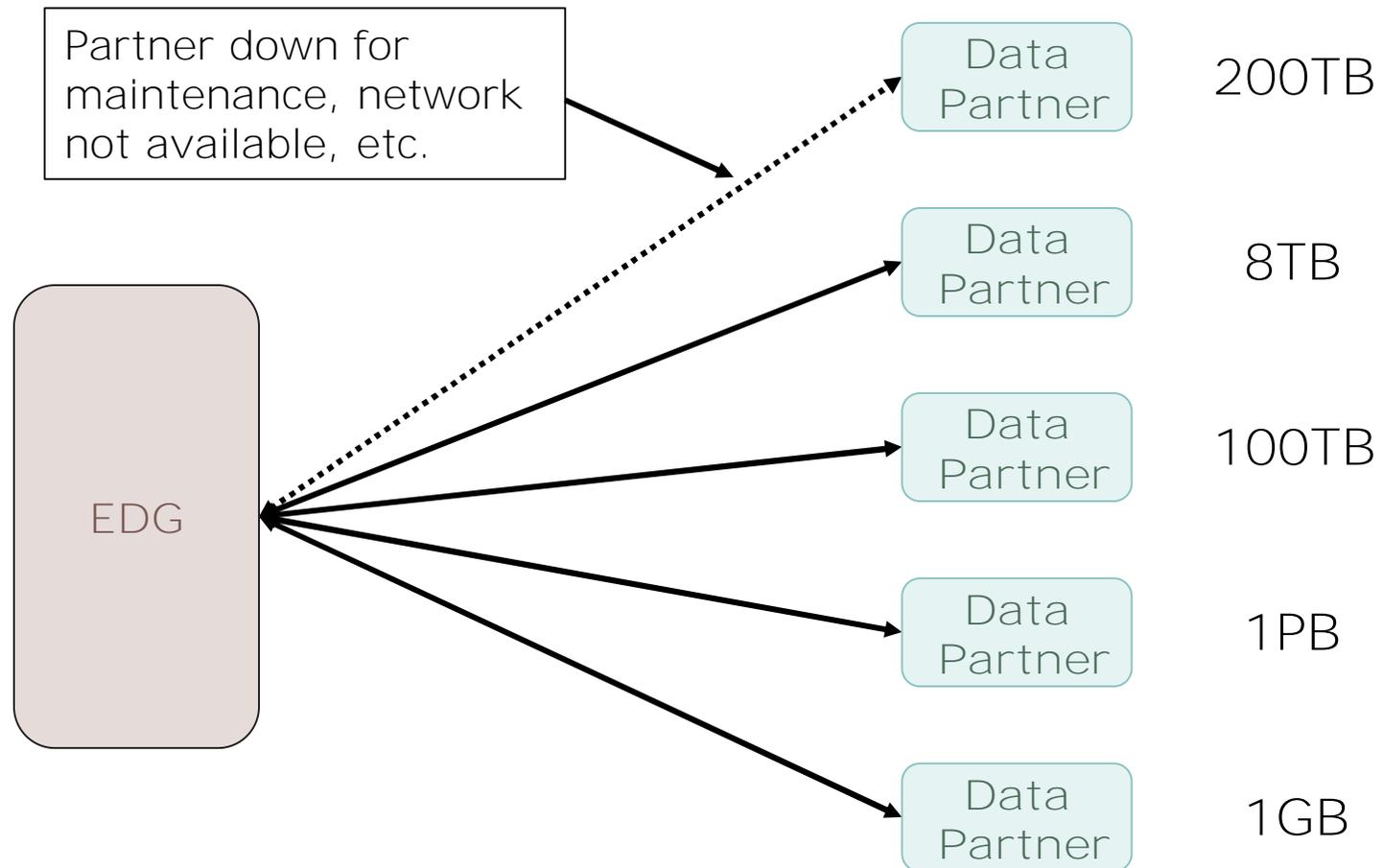
## Keyword Searches

## Numeric Searches

## Product Specific Attribute Searches

Extensibility to support varying, specialized inventory search capabilities. E.g spatial models

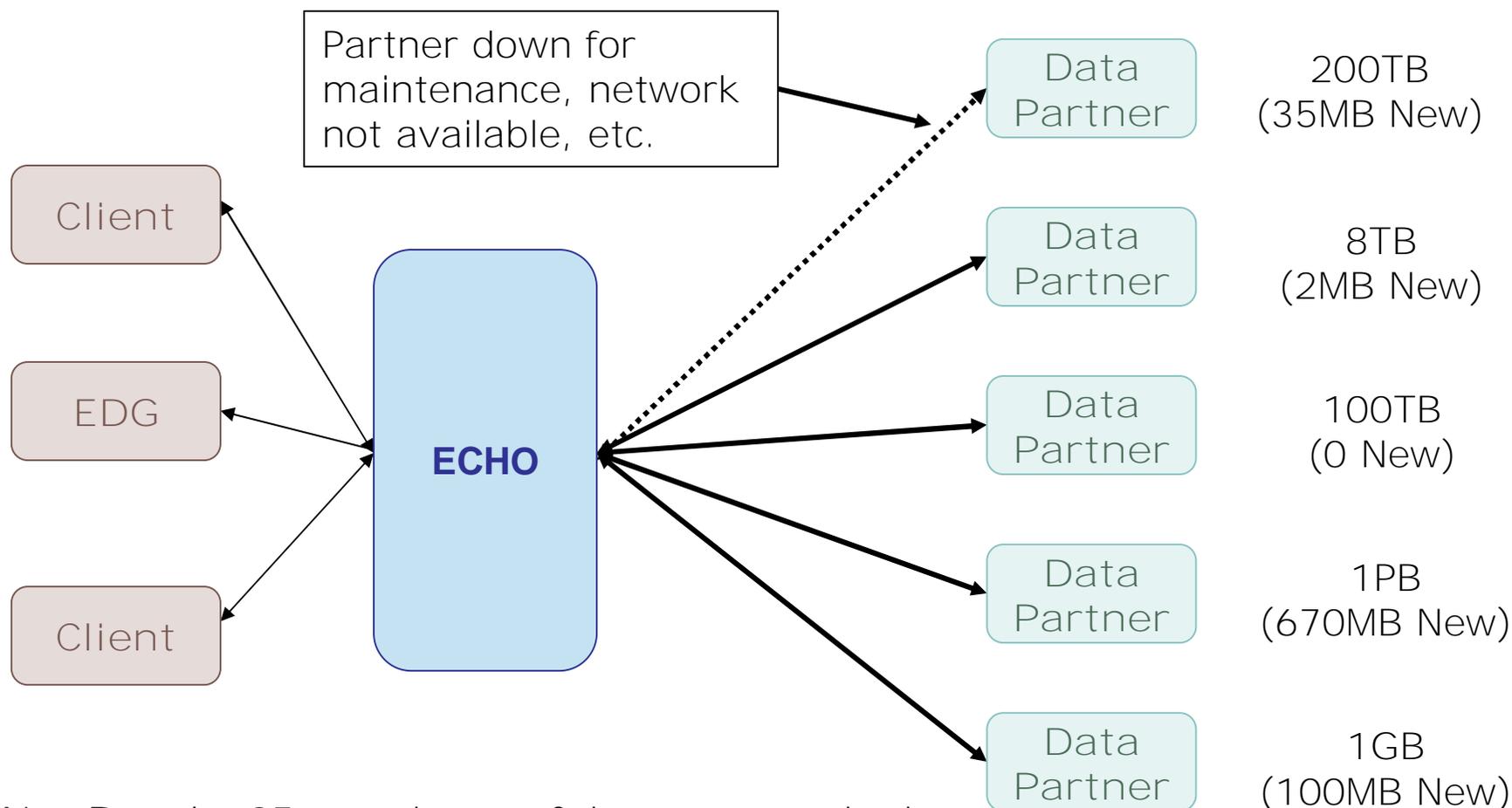
# Improved Availability: Old Architecture



Net Result: 200 terabytes of data not searched.  
Search time dependent on lowest common denominator.



# Improved Availability: ECHO Architecture



Net Result: 35 megabytes of data not searched.  
Search time dependent on ECHO only.



## Online Access or Order

---



Some Data Partners make the actual data available online. In this case, the OnlineAccessURL is included in the query result, and users can access the data without going through the order process provided in Order Entry Service.

If the data is not online, the Order Entry Service allows users to create, modify, quote, and submit orders.

The user can also check the status of the order if it is updated by the Data Partner.

One order consists of one or more provider orders. Each provider order represents the catalog items that a user wants from a particular Data Partner.

Each provider order consists of one or more order line items, with each line item describing the catalog item that is being requested including any requested options for that item.

- Catalog Items in ECHO are either granules or collections
- Only items that are orderable are Catalog Items



# Subscription Service

---

The subscription service automatically keeps users or clients informed of updated metadata

- ∅ Don't have to issue a query every night
- ∅ Only updated metadata is sent

The user can subscribe to the following options

- ∅ Collection metadata updates of one or all collections
- ∅ Granule metadata updates of a collection
- ∅ Both collection metadata and granule metadata updates of a collection

The metadata can be delivered by either FTP or e-mail

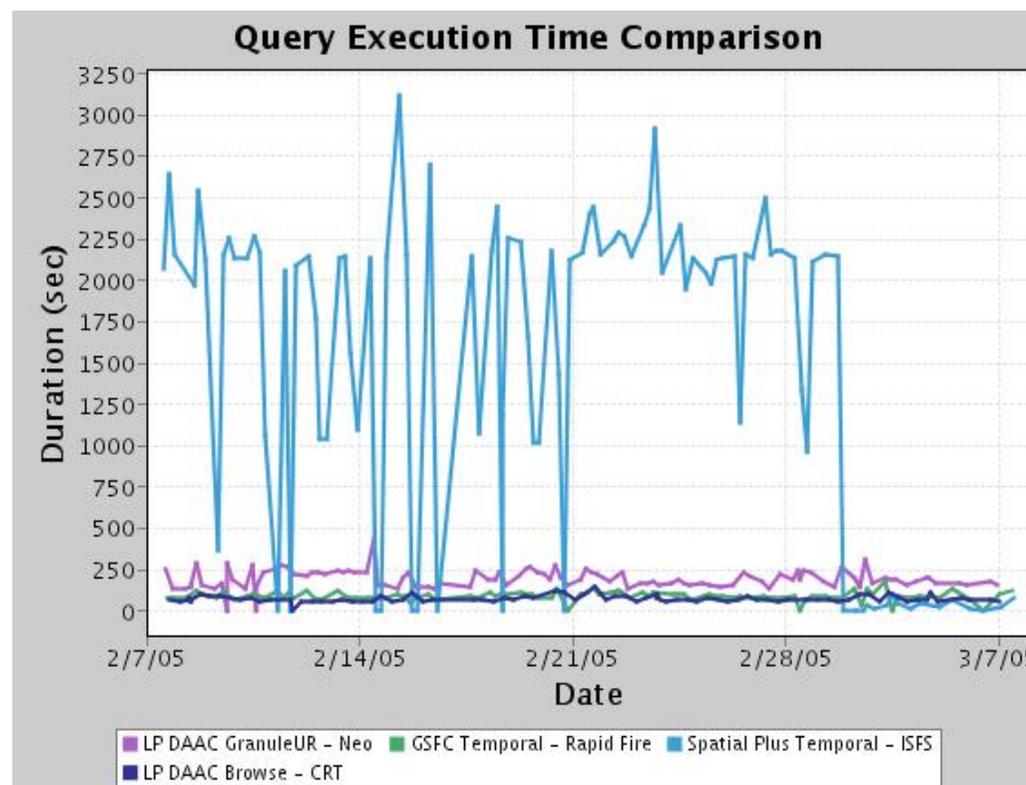
Must be a registered user to use this service

- ∅ Email address of user is used to report errors

The subscription service honors the Access Control Lists by only delivering the data that the registered user has access to

# Query Performance

	<u>5.5.2</u>	<u>5.5.3</u>
<b>GranuleUR</b>		
∅ # of Runs	447	13
∅ Min:	120	152
∅ Max:	439	205
∅ Mean:	185	178
<b>DataSetID, Browse Only</b>		
∅ # of Runs	350	13
∅ Min:	45	59
∅ Max:	378	108
∅ Mean:	83	77
<b>DataSetID, Temporal</b>		
∅ # of Runs	440	10
∅ Min:	40	71
∅ Max:	447	176
∅ Mean:	95	105
<b>DataSetID, Temporal, Spatial</b>		
∅ # of Runs	173	11
∅ Min:	311	15
∅ Max:	3575	98
∅ Mean:	1810	40





# System Availability

---



12/1/2004 - 3/1/2005

## Operational System Uptime

- Overall average: 98.5%
- December 2004: 96.4%
- January 2005: 99.6%
- February 2005: 99.4%

## Partner Test System Uptime

- Overall average: 99.8%
- December 2004: 99.8%
- January 2005: 99.8%
- February 2005: 99.7%



## WI ST



---

Warehouse Inventory Search Tool – general  
ECHO client based on EDG GUI

### WI ST 0.2

- ∅ ECHO 6.0 compatible
- ∅ June 1, 2005

### WI ST 0.3

- ∅ ECHO 7.0 compatible
- ∅ Sept 1, 2005

### Draft of EDG to WI ST transition plan

- ∅ Distributed to the DAAC Managers this week
- ∅ Telecons